

---

Termin: 22.1.2007

**Aufgabe 5.1:**

Konstruieren Sie für  $S = abcabbabcaa$  den Suffixbaum (nach dem Algorithmus von McCreight).

**Aufgabe 5.2:**

Konstruieren Sie für  $S = fitzefitzefatze$

- (a) das Suffix-Array und das inverse Suffix-Array mittels Verfeinerung,
- (b) das LCP-Array,
- (c) den Baum der LCP-Intervalle.

**Aufgabe 5.3:**

Ein Wort  $\alpha$  heißt *eindeutiges Teilwort* (*unique match*) von  $S$ , wenn es genau einmal in  $S$  vorkommt.

- (a) Wie sind die eindeutigen Teilwörter von  $S$  durch den Suffixbaum charakterisiert?
- (b) Entwickeln Sie aus der Charakterisierung einen Linearzeitalgorithmus zur Bestimmung des *kürzesten* eindeutigen Teilwortes von  $S$ .

**Aufgabe 5.4:**

Zeigen Sie, dass die supermaximalen Wiederholungen in einem Wort  $S$  wie folgt mit Hilfe des Suffix-Arrays  $A$  sowie des LCP-Arrays  $LCP$  von  $S$  charakterisiert werden können.

Das Wort  $\alpha$  der Länge  $\ell$  ist genau dann eine maximale Wiederholung in  $S$ , wenn es Indizes  $i < j$  mit folgenden Eigenschaften gibt:

1.  $LCP[i - 1] < \ell$ ,  $LCP[i] = LCP[i + 1] = \dots = LCP[j - 1] = \ell$ ,  $LCP[j] < \ell$ .
2. Die Zeichen  $S[A[i] - 1]$ ,  $S[A[i + 1] - 1]$ ,  $\dots$ ,  $S[A[j] - 1]$  sind paarweise verschieden.
3. An der Position  $A[i]$  von  $S$  kommt  $\alpha$  vor.

Entwickeln Sie mittels dieser Charakterisierung einen Linearzeitalgorithmus zur Bestimmung aller supermaximalen Wiederholungen in  $S$ .